Fórum:Testy2/Standardizace jako stanovení norem

Samotný výsledek konkrétního testu nemá žádnou vypovídací hodnotu o tom, jak si respondent stojí v porovnání s ostatními. Pouhý počet bodů nám neříká nic o tom, jestli student dosáhl nadprůměrného výsledku, nebo naopak zapadl do beznadějného podprůměru. Získá-li pak student ve dvou různých testech stejný počet bodů, může to znamenat v jednom testu vynikající výkon, zatímco v testu druhém pouze výkon průměrný. Teprve na základě porovnání dosaženého počtu bodů se standardy nebo výkony ostatních jsme schopni jednotlivce adekvátně posoudit. Standardizovanými testy se tedy snažíme o vyjádření výsledků jednotlivých respondentů buď vzhledem k výsledkům reprezentativního vzorku (typicky se jedná o stovky studentů) [1], nebo vzhledem ke kritériím – konkrétním znalostem, které absolvent kurzu musí mít.

Nejjednodušší metody standardizace jsou založeny na určení procenta respondentů, kteří dosáhli v daném testu horšího výsledku než daný student. Tento postup se používá často například při vyhodnocení přijímacího řízení. Ke každému bodovému zisku je přiřazeno *percentilové pořadí*, které zhruba uvádí, kolik procent respondentů dosáhlo výsledku horšího než testovaný uchazeč. Lze tak velmi snadno posoudit relativní pořadí konkrétního jedince v celé skupině respondentů.

Standardizaci testu – ve smyslu objektivizace hodnocení výsledku studenta v testu – lze rozdělit na tři přístupy

- **Relativní standardizace** je založena na analýze rozdělení získaných dat a porovnává výsledky respondentů mezi sebou.
- **Absolutní standardizace** je založena na dosažení konkrétních kritérií, tedy například toho, kolik správně zodpovězených otázek daný respondent vyprodukoval. Příkladem je stanovení hranice 70 % správně zodpovězených otázek pro úspěšné složení testu. [2]
- **Kombinovaná standardizace** je pak kombinací absolutní hranice mezi úspěšným a neúspěšným studentem (tzv. *pass mark*) a relativního rozdělení známek v pásmu úspěšnosti např. podle percentilů, směrodatné odchylky, apod.

Tab. 7.1 Schematický příklad ilustrující hodnocení studentů při relativní a absolutní standardizaci.
Studenti mají zodpovědět otázku: Co bylo příčinou druhé světové války?

Odpovědi studentů	Hodnocení při absolutní standardizaci	Hodnocení při relativní standardizaci
Student 1: Druhá světová válka byla vyvolána vpádem Hitlera do Polska.	Odpověď je správná.	Tato odpověď je horší než odpověď Studenta 2, ale lepší než odpověď Studenta 3.
Student 2: Druhá světová válka byla vyvolána mnoha faktory včetně hospodářské krize, obecné ekonomické situace, růstu nacionalismu a nevyřešených následků první světové války. Válka v Evropě začala německou invazí do Polska.	Odpověď je správná.	Tato odpověď je lepší než odpověď Studenta 1 a Studenta 3.
Student 3: Druhá světová válka byla vyvolána atentátem na Arcivévodu Ferdinanda.	Odpověď je chybná.	Tato odpověď je horší než odpověď Studenta 1 a Studenta 2.

Rozhodnutí, který typ standardizace pro konkrétní test použijeme, souvisí vždy s účelem testu.

Výhody a nevýhody jednotlivých typů standardizace

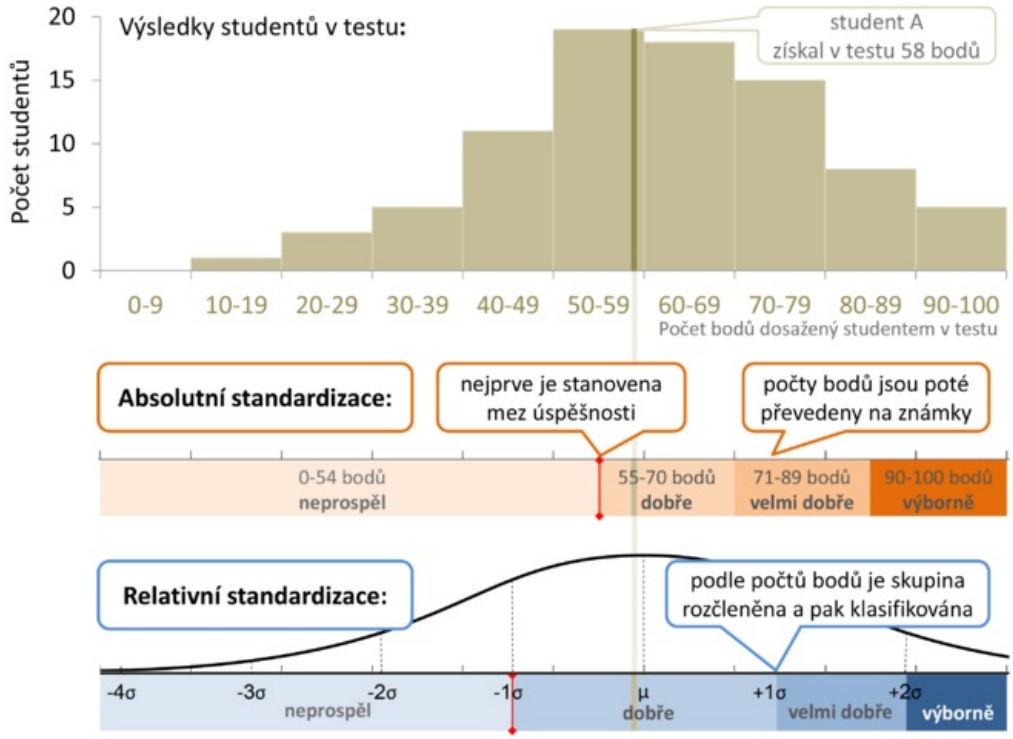
Relativní standardizace se neváže na obsah testu, ale hodnotí jednotlivé účastníky mezi sebou. Výhodou tedy je, že zabraňuje inflaci nejvyšších hodnocení, zřetelně odliší nejlepší studenty a není nutné individuálně standardizovat každý test zvlášť.

Mezi nevýhody relativního hodnocení patří kolísání kvality úspěšných studentů podle kvality dané skupiny. Zejména u menších skupin se tedy může stát, že uspějí i studenti s úrovní znalostí, která neodpovídá našim požadavkům. A obráceně, část studentů nemůže v testu uspět, ani kdyby látku uměli sebelépe. Hodnocení studentů podle relativní standardizace odrazuje od spolupráce a týmové práce, protože si studenti uvědomují, že si navzájem konkurují o omezený počet nejvyšších hodnocení. Snižuje to i motivaci studentů oslabením vztahu mezi jejich úsilím a výslednou známkou, protože ta závisí nejen na jejich vlastním výkonu, ale i na výkonu ostatních. Zvláště v menších a homogenních skupinách může relativní standardizace zveličit nepodstatné rozdíly. S ohledem na tato omezení bychom o užití relativního hodnocení měli uvažovat především ve velkých heterogenních skupinách, v nichž se nepředpokládá spolupráce.

Absolutní hodnocení závisí jen na tom, co se student naučil, nikoli na jeho pozici mezi ostatními. Jeho nevýhodou je nutnost stanovovat kritéria úspěchu pro každý test zvlášť. Musí být nastavena tak, aby rozlišovala mezi studenty, kteří danou oblast dostatečně zvládli, a těmi, jejichž znalosti či dovednosti nejsou dostatečné k dalšímu postupu.

Kombinovaná standardizace spojuje do jisté míry výhody absolutního hodnocení s kompetitivním aspektem hodnocení relativního. Studenti, kteří dosáhli absolutní hranice pro úspěšné složení zkoušky, jsou rozřazeni do skupin a podle dosažených bodů jsou jim přiděleny známky.

Níže je uvedeno schéma, které ilustruje rozdílný přístup k hodnocení při relativní a absolutní standardizaci.



Obr. 7.2 **Příklad hodnocení výsledku studenta v testu z pohledu absolutní a relativní standardizace**.

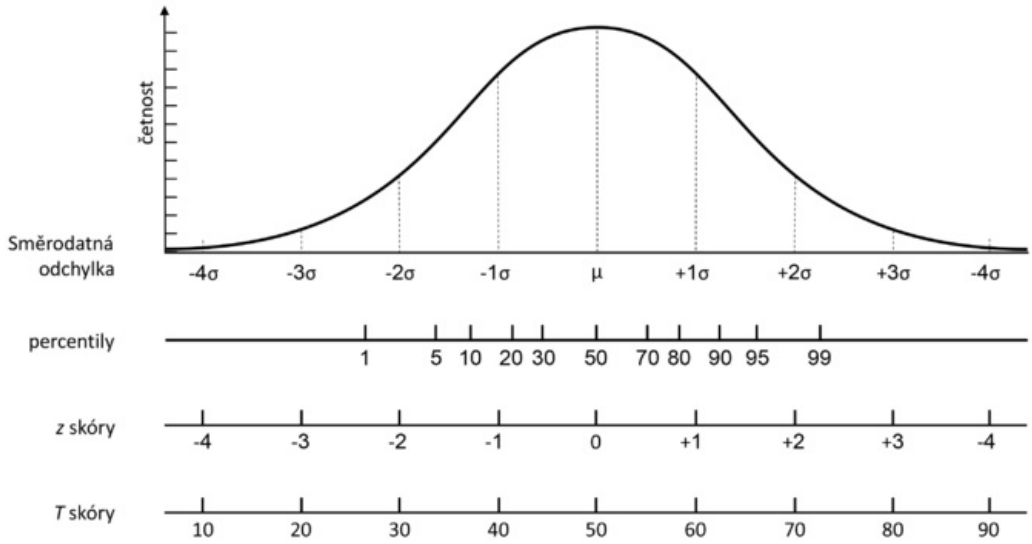
Absolutní standardizace srovnává výkon studenta s hranicí vědomostí vyžadovanou pro absolvování kurzu. Hranice se stanovují na základě expertního odhadu. Student A v testu uspěl, neboť dosáhl počtu bodů, který učitel považoval za minimum nezbytné pro absolvování zkoušky.

Relativní standardizace je založena na porovnávání výsledků studentů mezi sebou – skupina je rozčleněna podle dosaženého počtu bodů a oznámkována. Student A byl klasifikován známkou „dobře“. Relativní standardizaci můžeme provést až po testu, kdy jsou již známé výsledky studentů.

Relativní standardizace

Relativní standardizace je způsob vyhodnocení testu, při němž se výkon testovaného jedince porovnává s výkonem relevantní populace. To znamená, že se zjišťuje, zda zkoušený jedinec dosahuje lepších nebo horších výsledků než ostatní testovaní. Testům, při nichž se výkon testovaného posuzuje v relaci k ostatním, se anglicky říká *norm-referenced tests*, (NRT). Tento přístup k hodnocení výsledku jednotlivce v kontextu výkonu ostatních používají například zkoušky SAT, používané jako rozhodující kritérium pro přijetí na mnohé vysoké školy v USA. V našem prostředí je relativní standardizace porovnávající výkon studentů mezi sebou běžnou součástí přijímacích zkoušek či různých rozřazovacích testů.

Nevýhodou relativní standardizace je, že hodnocení jednotlivce nezávisí jen na jeho výkonu, ale i na výkonech ostatních studentů. Relativní standardizace je vhodná pro porovnávání výkonu velkých skupin a neměla by být používána ve skupinách menších než 40 studentů.



Obr. 7.3 Relativní standardizace porovnává výkon jednotlivce s ostatními zkoušenými. Celkové skóre se přitom převádí na odvozené hodnoty. K vyjádření studentova výsledku ve skupině lze použít některou z metod relativní standardizace testu:
Percentilová škála zhruba udává, jaké procento testované populace dosahuje horších výsledků než daný student.
z-škála popisuje, jak daleko (měřeno směrodatnou odchylkou dat) je výsledek daného studenta od průměru.
T-škála používá stejnou metriku, ale vyjadřuje ji na stovkové stupnici.

> 💡 **Tip: Samotné celkové skóre, resp. celkový počet bodů, může dávat zkreslující obraz o výsledku studenta. Chcete-li znát výsledek studenta vzhledem k testované skupině, použijte některou z metod relativní standardizace.**

Percentilová škála

Nejznámější metodou porovnávající vzájemně výkony testovaných je zobrazení jejich výkonů pomocí **percentilové škály**. K výsledku studenta se zjistí *percentil*, který zhruba říká, kolik procent studentů referenční skupiny mělo výsledek horší než daný student. Percentil tak přibližně určuje pořadí studenta přepočítané na interval 0 až 1 (resp. 0 - 100%).

Při výpočtu percentilu dosaženého studentem se spočítá počet studentů, kteří měli výsledek horší než daný student, přičte se polovina studentů, kteří měli výsledek stejný jako daný student, a určí se, jak velkou část tvoří tato skupina. Percentilové pořadí PR_i pro osobu s *i*-tým nejhorším celkovým skóre lze odvodit prostřednictvím vztahu:

$$PR_i = 100 \cdot \frac{N_i - \frac{n_i}{2}}{n},$$

kde N_i je kumulativní četnost u daného výsledku, n_i je četnost daného výsledku a n je počet testovaných žáků. Kumulativní četnost vyjadřuje počet studentů, kteří dosáhli daného nebo horšího výsledku.

Uvažujme, že chceme vypočítat percentilová pořadí 30 testovaných studentů, kteří dosáhli následujících výsledků:

1, 5, 5, 8, 8, 8, 8, 15, 15, 15, 15, 16, 16, 16, 21, 21, 23, 23, 23, 23, 24, 25, 27, 28, 28, 28, 28, 28, 28, 28

*Nejprve sestavíme tabulku četností, tedy k danému výsledku (celkovému počtu bodů) uvedeme **četnost** všech studentů se stejným bodovým výsledkem. Poté spočítáme **kumulativní četnosti** jako součet četností v daném řádku tabulky a všech předchozích řádcích. Nakonec dopočítáme pro každý řádek percentilové pořadí dle výše uvedeného vztahu.*

Uvažujeme-li studenta, který získal v didaktickém testu 21 bodů, pak je

- *jeho celkové skóre v pořadí 6. nejhorší, tedy* $i = 6,$
- *četnost daného výsledku je* $n_6 = 2,$
- *kumulativní četnost daného výsledku je* $N_6 = 16,$
- *počet testovaných žáků byl* $n = 30.$

Tedy percentilové pořadí pro studenta s 21 body je

Tab. 7.2 Výpočet percentilů

i	**Počet bodů z testu**	**Četnost** (n_i)	**Kumulativní četnost** (N_i)	**Percentilové pořadí** (PR_i)
1	1	1	1	1,67
2	5	2	3	6,67
3	8	4	7	16,67
4	15	4	11	30,00
5	16	3	14	41,67
6	**21**	**2**	**16**	**50,00**
7	23	4	20	60,00
8	24	1	21	68,33
9	25	1	22	71,67
10	27	1	23	75,00
11	28	7	30	88,33

$$PR_6 = 100 \cdot \frac{N_6 - \frac{n_6}{2}}{n} = 100 \cdot \frac{16 - \frac{2}{2}}{30} = 50,$$

jinými slovy mezi 100 studenty by se náš student s 21 body umístil zhruba na 50. místě. Zhruba 50 % studentů testované (referenční) skupiny dosáhlo horšího výsledku než student s 21 body.

z-skóry

Další metodou standardizace výsledku studenta je výpočet jeho z-skóru. **z-skór** daného studenta ukazuje, **nakolik je jeho výsledek nad nebo pod průměrem** (měřeno v jednotkách **směrodatné odchylky**). z-skór je tedy počítán jednoduše jako rozdíl studentova hrubého skóru X a průměru celé skupiny $\bar{X}$, vydělený směrodatnou odchylkou SD:

$$z = \frac{X - \bar{X}}{SD}.$$

Pokud například studenti dosáhli v testu průměrně $\bar{X} = 50$ bodů a směrodatná odchylka byla $SD = 20$, pak student, který měl 30 bodů, má z-skór roven $(30 - 50)/20 = -1.$ To znamená, že studentův výsledek je jednu směrodatnou odchylku pod celkovým průměrem. Pokud se celkové skóre řídí přibližně normálním rozdělením, pak platí, že přibližně 68 % studentů má výsledný počet bodů v rozmezí ±1 směrodatná odchylka od průměru, celkem 95 % studentů má výsledek testu v rozmezí ±2 směrodatné odchylky a až 99,75 % studentů v rozmezí ±3 směrodatné odchylky od průměru. Pokud je tedy studentův z-skór roven −1, znamená to, že celkem asi (100 − 68)/2 = 16 % studentů má výsledek testu nižší než zmíněný student.

Pomocí z-skóru může vyučující snadno identifikovat žáky výtečné (z > 2) a naopak velmi slabé (z < −2). Snadno může také porovnat studentovy výsledky v různých částech testu.

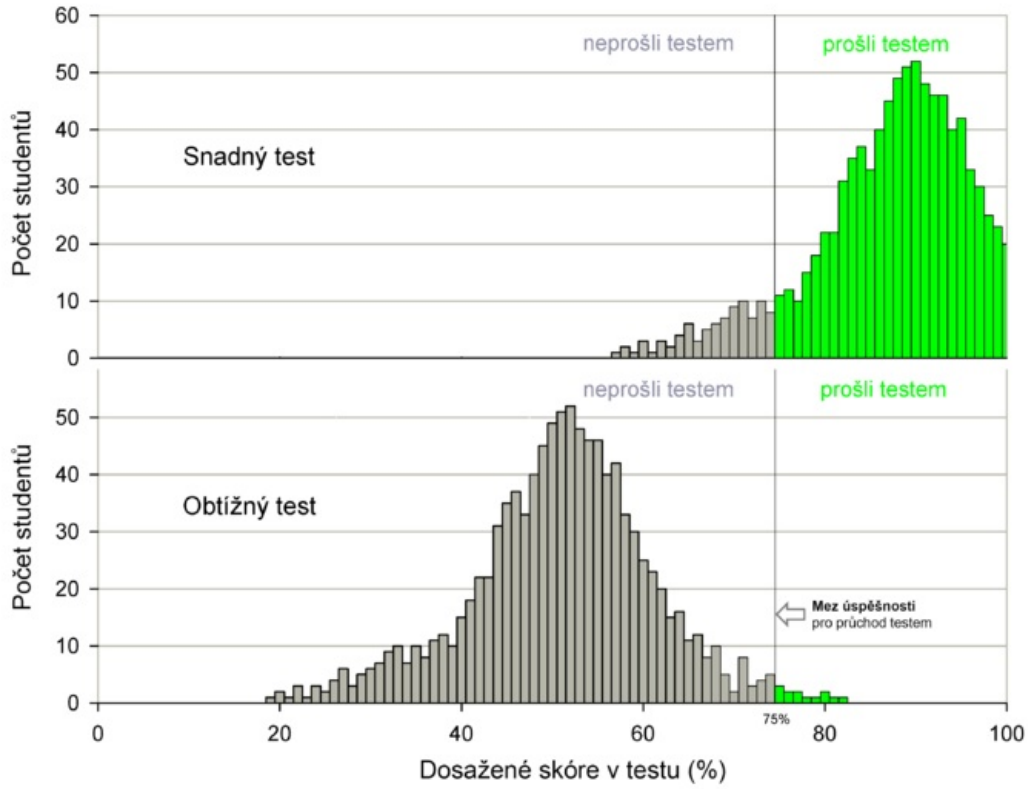
Podrobnější rozbor dalších metod standardizace (C-škála, škála „stanin“) je k dispozici například v publikaci autorů Jeřábka a Bílka s názvem Teorie a praxe tvorby didaktických testů. **Cite error: Invalid <ref> tag; invalid names, e.g. too many**

Absolutní standardizace

Absolutní standardizace je způsob hodnocení testu, při němž se výkon studenta porovnává s absolutními kritérii – s požadavkem na nabytí vědomostí nebo dovedností, které musí mít, aby bylo možno považovat jeho znalost za dostatečnou pro úspěšné absolvování testu (a kurzu). Kritériem se přitom myslí dosažení konkrétní vědomosti a schopnosti, nikoliv dosažení určitého počtu bodů v testu. Např. stanovíme, že po absolvování kurzu první pomoci by měl frekventant znát doporučení týkající se kardiopulmonální resuscitace, jinak musí kurz absolvovat znovu. Jiným příkladem absolutně standardizovaného testu je test v autoškole: je důležité nevypouštět do ulic řidiče, kteří neznají základní předpisy, a to ani v případě, že by se řadili mezi relativně lepší uchazeče. Absolutně standardizované testy se v angličtině nazývají *criterion-referenced tests* (CRT) a používají se například při národních licenčních zkouškách zdravotních sester v USA (National Council Licensure Examination (NCLEX)).

V případě absolutního hodnocení testu je třeba správně zvolit hranici mezi úspěšným a neúspěšným studentem, tedy rozhraní mezi studenty, kteří danou oblast zvládli dostatečně, a těmi, kteří ji dostatečně nezvládli. Ke stanovení této hranice se bohužel občas používají intuitivní či „tradiční“, víceméně arbitrární procentuální meze

(50 %, 75 % apod.) bez hlubšího zdůvodnění. Přitom některé otázky v testu mohu mít zásadní význam a jiné mohou být jen okrajové.



Obr. 7.4 Dvojice grafů ukazuje, jaké důsledky by mohlo mít nastavení meze úspěšnosti v testu odhadem „od oka“. V případě snadného testu (obrázek nahoře) uspějí při nastavení limitu na 75 % téměř všichni. Stejný limit použitý pro jiný, obtížnější test (obrázek dole) způsobí, že testem projde jen pár žáků.

Abychom se vyhnuli problémům se špatně stanovenou mezí úspěšnosti, je třeba **kvalifikovaně posoudit obtížnost testu**, tedy provést jeho **absolutní standardizaci**. Dobře k tomu poslouží např. Angoffova či Ebelova metoda, které jsou popsány dále.

Existuje celá řada metod pro absolutní standardizaci různých typů hodnocení studentů. Jejich přehled nalezne čtenář například v obsáhlém díle Handbook of Test Development [3].

Většina metod vychází z **expertního posudku** relevantních odborníků, který se může zaměřit buď na jednotlivé položky testu (tedy jak která položka přispívá k rozlišení mezi úspěšným a neúspěšným studentem), nebo naopak na testovanou populaci (tedy zda test umí rozlišit mezi vhodnými a nevhodnými kandidáty).

Zaměřme se nyní podrobněji na dvě metody standardizace – Angoffovu a Ebelovu metodu. K praktickému provedení je v obou případech vhodné zajistit 6–8 odborníků v testovaném oboru.

Angoffova metoda

Patrně nejznámější z položkových metod standardizace je metoda podle **Angoffa**, resp. její modifikace podle Hambletona a Plakeové [4]. Principem metody je expertní odhad hraničního počtu bodů nutného k úspěšnému absolvování testu nebo úlohy. Pracuje se s hypotetickým tzv. minimálně kompetentním studentem. *Minimálně kompetentní student* je takový, který právě splňuje minimální požadavky kladené v daném oboru. Jinými slovy, je to nejslabší student, který by ještě měl testem projít.

Test (nebo položku) posuzuje skupina expertů (obvykle tým učitelů daného oboru či kurzu, kteří zodpovídají za přípravu a hodnocení testů, *examination committee*) a každý z nich u každé položky odhaduje, jaké procento minimálně kompetentních studentů by na danou otázku odpovědělo správně. Experti pracují samostatně, aby se vzájemně neovlivňovali. Výsledky se zapíší do tabulky, v níž je na každém řádku určitá položka z testu a každý sloupec tvoří odhady jednoho experta.

Tab. 7.3 Tabulka expertních odhadů pravděpodobnosti správného zodpovězení otázky minimálně kompetentním studentem

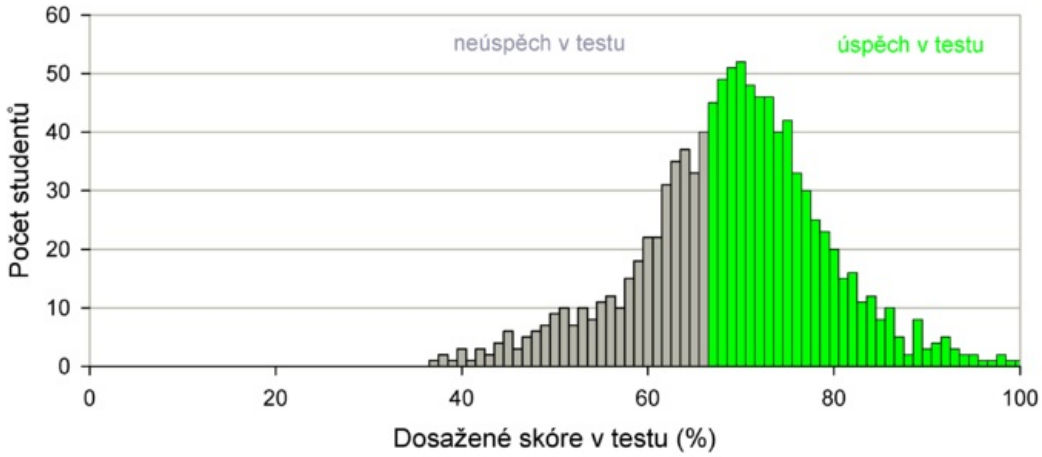
Číslo položky	Expert 1	Expert 2	Expert 3	Průměr
1	0,75	0,60	0,60	0,65
2	0,70	0,50	0,60	0,60
3	0,80	0,60	0,70	0,70

4	0,70	0,60	0,70	0,70
…	…	…	…	…
Průměr				**0,66, tj. 66 %**

Pokud by byly v testu otázky typu ANO/NE, byl by samozřejmě nejnižší možný odhad úspěšnosti 0,5, neboť i student, který odpověď nezná, má 50 % pravděpodobnost odpovědět správně. Analogicky u výběrové otázky s pěti možnostmi a jedinou správnou odpovědí bude minimum 0,2.

Po vyplnění tabulky se obvykle posuzuje, zda se experti ve svých odhadech shodli. Položky, v nichž je rozptyl odhadů velký, je třeba prodiskutovat; často se odhalí nejednoznačná formulace či jiný problém.

Na závěr se vypočte průměr všech odhadů v tabulce. Tento průměr říká, kolik procent z celkového možného počtu bodů by měl dosáhnout minimálně kompetentní student. Jinými slovy tento průměr udává mez úspěšnosti pro daný test – tedy hranici mezi „prošel“ a „neprošel“.



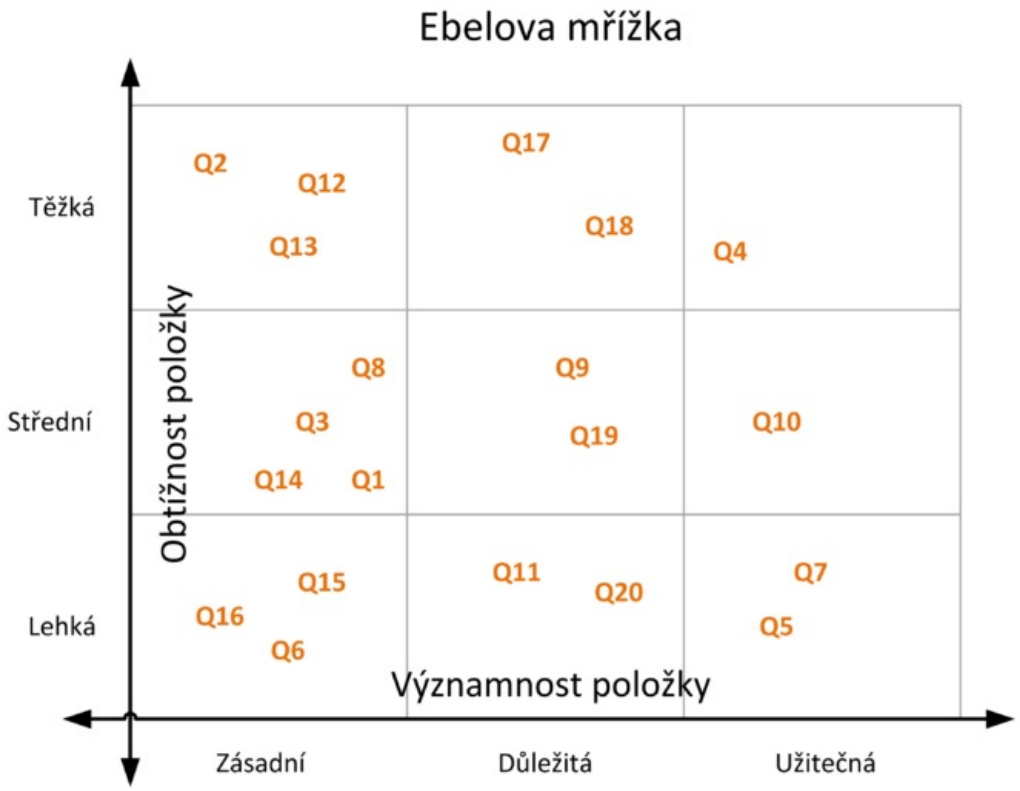
Obr. 7.5 Mez úspěšnosti stanovená pomocí standardizační metody rozdělí soubor testovaných na úspěšné a neúspěšné. Standardizaci umožňují a podporují i některé testovací programy, například Rogo, z nějž můžete obdržet právě takovýto graf.

Pro úspěšné použití této metody je nutné, aby zúčastnění experti měli dostatek zkušeností v dané oblasti a poměrně přesnou představu o tom, co studenti v daném kurzu musí zvládnout. Experti si tedy musí umět představit, co minimálně kompetentní student umí, resp. by měl umět. Vzorovou tabulku ve formátu MS Excel, která je navržena v souladu s Angoffovou metodou standardizace najdete v příloze (http://is.muni.cz/www/98951/standardizace-angoffova-metoda.xls).

Ebelova metoda

Ebelova metoda má dvě modifikace. **Tradiční Ebelova metoda** slouží pro stanovení expertního odhadu minimálního výsledku, kterého by měl student dosáhnout, aby ještě prošel testem [5], [6]. **Modifikovaná Ebelova metoda** slouží pro přípravu obsahově validního testu [7], [8], [9].

V tradiční metodě podle **Ebela** nejprve experti roztřídí jednotlivé otázky do skupin podle dvou kritérií: **významu** a **obtížnosti**. Škála významu (relevance) položky jde od „zásadní“, přes „důležitá“ a „užitečná“ až k „irelevantní“. Položky označené experty jako „irelevantní“ se z dalšího použití vyloučí. Škála obtížnosti je trojstupňová: „těžká“, „střední“ a „snadná“ [10]. Zařazením položek do kategorií podle obou kritérií vznikne **Ebelova mřížka**:



Obr. 7.6 Ebelova mřížka

Předem bývá dohodnuto, že položky, u nichž se nedosáhne definované úrovně shody expertů (např. 80 %), budou diskutovány a případně vyloučeny. Po takovémto rozdělení položek odhadnou experti, jakou část otázek z každé kategorie by měl správně zodpovědět minimálně kompetentní student. Součiny těchto proporcí a počtů otázek v každé kategorii se poté sečtou a toto číslo se vydělí celkovým počtem otázek: výsledkem je hledaná hranice úspěšnosti.

Tab. 7.4 Ebelova metoda – krok 1
Experti rozdělí otázky podle dvou kritérií – **význam** a **obtížnost**

	Zásadní	**Důležitá**	**Užitečná**
Těžká	**3**	**2**	**1**
Střední	**4**	**2**	**1**
Lehká	**3**	**2**	**2**

Tab. 7.5 Ebelova metoda – krok 2
Experti odhadnou úspěšnost zodpovězení otázkek z každé kategorie minimálně kompetentním studentem

	Zásadní	**Důležitá**	**Užitečná**
Těžká	**50 %**	**50 %**	**30 %**
Střední	**70 %**	**70 %**	**50 %**
Lehká	**90 %**	**80 %**	**60 %**

Tab. 7.6 Ebelova metoda – krok 3
Vypočtou se parametry pro jednotlivé kategorie: **počet otázek · odhad úspěšnosti MKS**

	Zásadní	**Důležitá**	**Užitečná**
Těžká	**3 · 0,5 = 1,5**	**2 · 0,5 = 1,0**	**1 · 0,3 = 0,3**

Střední	**4 · 0,7 = 2,8**	**2 · 0,7 = 1,4**	**1 · 0,5 = 0,5**
Lehká	**3 · 0,9 = 0,3**	**2 · 0,8 = 1,6**	**2 · 0,6 = 1,2**

Nakonec vypočteme hranici úspěšnosti: sečteme odhady pro jednotlivé kategorie z předešlé tabulky a výsledek vydělíme celkovým počtem otázek. V našem případě tedy 13 : 20 = 0,65, tj. za úspěšné absolvování testu považujeme získání nejméně 65 % z celkového počtu bodů.

Stejný výpočet můžeme provést i v tabulce:

Tab. 7.7 Tabulka pro výpočet meze úspěšnosti

Obtížnost	Důležitost (Relevance)	Počet otázek (n)	Proporce (P)	Součin (n · P)
Těžká	Zásadní	3	0,50	1,5
Střední	Zásadní	4	0,70	2,8
Lehká	Zásadní	3	0,90	2,7
Těžká	Důležitá	2	0,50	1,0
Střední	Důležitá	2	0,70	1,4
Lehká	Důležitá	2	0,80	1,6
Těžká	Užitečná	1	0,30	0,3
Střední	Užitečná	1	0,50	0,5
Lehká	Užitečná	2	0,60	1,2
Celkem		**20**		**13,0**
Hranice úspěšnosti				13 : 20 = 0,65, tj. **65 %**

Uvedený příklad je jen ilustrativní, v praxi bychom měli pracovat s většími počty otázek.

O významu Ebelovy metody svědčí i to, že Ebelova mřížka je součástí některých testovacích programů. Konkrétně je přímo zahrnuta v testovacím programu Rogo, o němž budeme podrobně mluvit v části věnované realizaci testů.

Reference

1. CHRÁSKA, Miroslav. *Metody pedagogického výzkumu.* 1. vydání. Praha : Grada Publishing a.s., 2007. 265 s. ISBN 80-247-1369-1.
2. BOURSICOT, Katharine. *Introduction to standard setting.* London : St. George's University, 2011.
3. DOWNING, Steven M a Thomas M HALADYNA. *Handbook of test development.* 1. vydání. Mahwah : Lawrence Erlbaum Associates, 2006. 778 s. ISBN 9780805852653.
4. HAMBELTON, Ronald K a Barbara S PLAKE. Using an extended Angoff procedure to set standards on complex performance assessments. *Applied measurement in education.* 1995, roč. 8, vol. 8, no. 1, s. 41-55, ISSN 0895-7347 (Print), 1532-4818 (Online). DOI: 10.1207/s15324818ame0801_4 (http://dx.doi.org/10.1207%2Fs15324818ame0801_4).
5. CANTOR, Jeffrey A. A Validation of Ebel's Method for Performance Standard Setting through its Application with Comparison Approaches to a Selected Criterion-Referenced Test. *Educational and Psychological Measurement.* 1989, roč. 49, vol. 49, no. 3, s. 709-721, ISSN (Print) 0013-1644; (Online) ISSN: 1552-3888.
6. VIOLATO, Claudio, Anthony MARINI a Curtis LEE. A validity study of expert judgement procedures for setting cutoff scores on high stakes credentialing examinations using cluster analysis. *Evaluation and the Health Professions* [online]. 2003, roč. 26, vol. 26, no. 1, s. 59-72, dostupné také z <http://www.internationalgme.org/Resources/Pubs/Validity%20Cutoff%20Scores%20-%20Violato.pdf>. ISSN (Print) 0163-2787; (Online) 1552-3918. PMID: 22973420 (https://www.ncbi.nlm.nih.gov/pubmed/22973420).DOI: 10.1177/0163278702250082 (http://dx.doi.org/10.1177%2F0163278702250082).
7. AZIZ, Saman. *A Modified Ebel Standard Setting Method for a Medical School Clinical Skills Assessment.* Chicago : University of Illinois, 2005. 162 s.

8. BUTTERWICK, D. J, D.M PASKEVICH a A.L VALLEVAND, et al. Development of content-valid technical skill assessment instruments for athletic taping skills. *Journal of Allied Health.* 2006, roč. 35, vol. 35, no. 3, s. 149-157, ISSN (Print) 0090-7421, (Online) 1945-404X. PMID: 17036669 (https://www.ncbi.nlm.nih.gov/pubmed/17036669).
9. VIOLATO, Claudio, Lanree SALAMI a Sylvia MUIZNIEKS. Certification Examinations for Massage Therapists: A Psychometric Analysis. *Journal of Manipulative Physiological Therapeutics* [online]. 2002, roč. 25, vol. 25, no. 2, s. 111-115, dostupné také z <http://www.jmptonline.org/article/S0161-4754(02)70455-7/fulltext>. ISSN 0161-4754. DOI: 10.1067/mmt.2002.121413 (http://dx.doi.org/10.1067%2Fmmt.2002.121413).
10. LAFAVE, M, L KATZ a D.J BUTTERWICK. Development of a content-valid standardized orthopedic assessment tool (SOAT). *Advances in health sciences education : theory and practice.* 2008, roč. 13, vol. 13, no. 4, s. 397-406, ISSN (Print) 1382-4996, (Online) 1573-1677. PMID: 17203268 (https://www.ncbi.nlm.nih.gov/pubmed/17203268).